SEQUENCE LISTING

<110> Tang, Jordan J.N.
Hong, Lin
Ghosh, Arun K.

<120> Inhibitors of Memapsin 2 and Use Thereof

<130> OMRF 182

<140> Not Yet Assigned
<141> 2000-06-27

<150> 60/141,363
<151> 1999-06-28

<150> 60/168,060
<151> 1999-11-30

<150> 60/177,836
<151> 2000-01-25

<150> 60/178,368
<151> 2000-01-27

<150> 60/210,292
<151> 2000-06-08

<160> 31

<170> PatentIn Ver. 2.1

<210> 1
<211> 3252
<212> DNA
<213> Homo sapiens

<400> 1
gcgggagtgc tgcctgccca cggcacccag cacggcatcc ggctgcccct gcgcagcggc 60
ctgggggggcg ccccccctggg gctgcggctg ccccgggaga ccgacgaaga gcccgaggag 120
cccggccgga ggggcagctt tgtggagatg gtggacaacc tgaggggcaa gtcggggcag 180
ggctactacg tggagatgac cgtgggcagc cccccgcaga cgctcaacat cctggtggat 240
acaggcagca gtaactttgc agtgggtgct gccccccacc ccttcctgca tcgctactac 300
cagaggcagc tgtccagcac ataccgggac ctccggaagg gtgtgtatgt gccctacacc 360
cagggcaagt gggaagggga gctgggcacc gacctggtaa gcatccccca tggccccaac 420
gtcactgtgc gtgccaacat tgctgccatc actgaatcag acaagttctt catcaacggc 480
tccaactggg aaggcatcct ggggctggcc tatgctgaga ttgccaggcc tgacgactcc 540
ctggagcctt tctttgactc tctggtaaag cagacccacg ttcccaacct cttctccctg 600

cagctttgtg gtgctggctt ccccctcaac cagtctgaag tgctggcctc tgtcggaggg 660
agcatgatca ttggaggtat cgaccactcg ctgtacacag gcagtctctg gtatacaccc 720
atccggcggg agtggtatta tgaggtgatc attgtgcggg tggagatcaa tggacaggat 780
ctgaaaatgg actgcaagga gtacaactat gacaagagca ttgtggacag tggcaccacc 840
aaccttcgtt tgcccaagaa agtgtttgaa gctgcagtca aatccatcaa ggcagcctcc 900
tccacggaga agttccctga tggtttctgg ctaggagagc agctggtgtg ctggcaagca 960
ggcaccaccc cttggaacat tttcccagtc atctcactct acctaatggg tgaggttacc 1020
aaccagtcct tccgcatcac catccttccg cagcaatacc tgcggccagt ggaagatgtg 1080
gccacgtccc aagacgactg ttacaagttt gccatctcac agtcatccac gggcactgtt 1140
atgggagctg ttatcatgga gggcttctac gttgtctttg atcgggcccg aaaacgaatt 1200
ggctttgctg tcagcgcttg ccatgtgcac gatgagttca ggacggcagc ggtggaaggc 1260
ccttttgtca ccttggacat ggaagactgt ggctacaaca ttccacagac agatgagtca 1320
accctcatga ccatagccta tgtcatggct gccatctgcg ccctcttcat gctgccactc 1380
tgcctcatgg tgtgtcagtg gcgctgcctc cgctgcctgc gccagcagca tgatgacttt 1440
gctgatgaca tctccctgct gaagtgagga ggcccatggg cagaagatag agattcccct 1500
ggaccacacc tccgtggttc actttggtca caagtaggag acacagatgg cacctgtggc 1560
cagagcacct caggaccctc cccacccacc aaatgcctct gccttgatgg agaaggaaaa 1620
ggctggcaag gtgggttcca gggactgtac ctgtaggaaa cagaaaagag aagaaagaag 1680
cactctgctg gcgggaatac tcttggtcac ctcaaattta agtcgggaaa ttctgctgct 1740
tgaaacttca gccctgaacc tttgtccacc attcctttaa attctccaac ccaaagtatt 1800
cttcttttct tagtttcaga agtactggca tcacacgcag gttaccttgg cgtgtgtccc 1860
tgtggtaccc tggcagagaa gagaccaagc ttgtttccct gctggccaaa gtcagtagga 1920
gaggatgcac agtttgctat ttgctttaga gacagggact gtataaacaa gcctaacatt 1980
ggtgcaaaga ttgcctcttg aattaaaaaa aaactagatt gactatttat acaaatgggg 2040
gcggctggaa agaggagaag gagagggagt acaaagacag ggaatagtgg gatcaaagct 2100
aggaaaggca gaaacacaac cactcaccag tcctagtttt agacctcatc tccaagatag 2160
catcccatct cagaagatgg gtgttgtttt caatgttttc ttttctgtgg ttgcagcctg 2220
accaaaagtg agatgggaag ggcttatcta gccaaagagc tctttttag ctctcttaaa 2280
tgaagtgccc actaagaagt tccacttaac acatgaattt ctgccatatt aatttcattg 2340
tctctatctg aaccaccctt tattctacat atgataggca gcactgaaat atcctaaccc 2400
cctaagctcc aggtgccctg tgggagagca actggactat agcagggctg ggctctgtct 2460
tcctggtcat aggctcactc tttcccccaa atcttcctct ggagctttgc agccaaggtg 2520
ctaaaaggaa taggtaggag acctcttcta tctaatcctt aaaagcataa tgttgaacat 2580
tcattcaaca gctgatgccc tataacccct gcctggattt cttcctatta ggctataaga 2640
agtagcaaga tctttacata attcagagtg gtttcattgc cttcctaccc tctctaatgg 2700
cccctccatt tatttgacta aagcatcrca cagtggcact agcattatac caagagtatg 2760
agaaatacag tgctttatgg ctctaacatt actgccttca gtatcaaggc tgcctggaga 2820
aaggatggca gcctcagggc ttccttatgt cctccaccac aagagctcct tgatgaaggt 2880
catcttttтc ccctatcctg ttcttcccct ccccgctcct aatggtacgt gggtacccag 2940
gctggttctt gggctaggta gtggggacca agttcattac ctccctatca gttctagcat 3000
agtaaactac ggtaccagtg ttagtgggaa gagctgggtt ttcctagtat acccactgca 3060
tcctactcct acctggtcaa cccgctgctt ccaggtatgg gacctgctaa gtgtggaatt 3120
acctgataag ggagagggaa atacaaggag ggcctctggt gttcctggcc tcagccagct 3180
gcccmcaagc cataaaccaa taaamcaaga atactgagtc taaaaaaaaa aaaaaaaaaa 3240
aaaaaaaaaa aa 3252

<210> 2

<211> 488
<212> PRT
<213> Homo sapiens

<220>
<223> Purified Memapsin 2

<220>
<223> Amino Acids 28-48 are remnant putative propeptide residues

<220>
<223> Amino Acids 58-61, 78, 80, 82-83, 116, 118-121, 156, 166, 174, 246, 274, 276, 278-281, 283, and 376-377 are residues in contact with the OM99-2 inhibitor

<220>
<223> Amino acids 54-57, 61-68, 73-80, 86-89, 109-111, 113-118, 123-134, 143-154, 165-168, 198-202, and 220-224 are N-lobe Beta Strands

<220>
<223> Amino Acids 184-191 and 210-217 are N-lobe Helices

<220>
<223> Amino acids 237-240, 247-249, 251-256, 259-260, 273-275, 282-285, 316-318, 331-336, 342-348, 354-357, 366-370, 372-375, 380-383, 390-395, 400-405, and 418-420 are C-lobe Beta Strands

<220>
<223> Amino Acids 286-299, 307-310, 350-353, 384-387, and 427-431 are C-lobe Helices

<400> 2

```
Ala Gly Val Leu Pro Ala His Gly Thr Gln His Gly Ile Arg Leu Pro
  1               5                  10                  15

Leu Arg Ser Gly Leu Gly Gly Ala Pro Leu Gly Leu Arg Leu Pro Arg
             20                  25                  30

Glu Thr Asp Glu Glu Pro Glu Glu Pro Gly Arg Arg Gly Ser Phe Val
         35                  40                  45

Glu Met Val Asp Asn Leu Arg Gly Lys Ser Gly Gln Gly Tyr Tyr Val
     50                  55                  60
```

Glu Met Thr Val Gly Ser Pro Pro Gln Thr Leu Asn Ile Leu Val Asp
65 70 75 80

Thr Gly Ser Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu
85 90 95

His Arg Tyr Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg
100 105 110

Lys Gly Val Tyr Val Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu
115 120 125

Gly Thr Asp Leu Val Ser Ile Pro His Gly Pro Asn Val Thr Val Arg
130 135 140

Ala Asn Ile Ala Ala Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly
145 150 155 160

Ser Asn Trp Glu Gly Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg
165 170 175

Pro Asp Asp Ser Leu Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr
180 185 190

His Val Pro Asn Leu Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro
195 200 205

Leu Asn Gln Ser Glu Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile
210 215 220

Gly Gly Ile Asp His Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro
225 230 235 240

Ile Arg Arg Glu Trp Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile
245 250 255

Asn Gly Gln Asp Leu Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys
260 265 270

Ser Ile Val Asp Ser Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val
275 280 285

Phe Glu Ala Ala Val Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys
290 295 300

Phe Pro Asp Gly Phe Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala
305 310 315 320

Gly Thr Thr Pro Trp Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met
325 330 335

Gly Glu Val Thr Asn Gln Ser Phe Arg Ile Thr Ile Leu Pro Gln Gln
340 345 350

Tyr Leu Arg Pro Val Glu Asp Val Ala Thr Ser Gln Asp Asp Cys Tyr
355 360 365

Lys Phe Ala Ile Ser Gln Ser Ser Thr Gly Thr Val Met Gly Ala Val
370 375 380

Ile Met Glu Gly Phe Tyr Val Val Phe Asp Arg Ala Arg Lys Arg Ile
385 390 395 400

Gly Phe Ala Val Ser Ala Cys His Val His Asp Glu Phe Arg Thr Ala
405 410 415

Ala Val Glu Gly Pro Phe Val Thr Leu Asp Met Glu Asp Cys Gly Tyr
420 425 430

Asn Ile Pro Gln Thr Asp Glu Ser Thr Leu Met Thr Ile Ala Tyr Val
435 440 445

Met Ala Ala Ile Cys Ala Leu Phe Met Leu Pro Leu Cys Leu Met Val
450 455 460

Cys Gln Trp Arg Cys Leu Arg Cys Leu Arg Gln Gln His Asp Asp Phe
465 470 475 480

Ala Asp Asp Ile Ser Leu Leu Lys
485

<210> 3
<211> 503
<212> PRT
<213> Homo sapiens

<220>
<223> Pro-memapsin 2

<220>
<223> Amino Acids 1-15 are vector-derived residues

<220>
<223> Amino Acids 16-64 are a putative pro peptide

<220>
<223> Amino Acids 1-13 are the T7 promoter

<220>
<223> Amino Acids 16-456 are Pro-memapsin 2-T1

<220>
<223> Amino Acids 16-421 are Promemapsin 2-T2

<400> 3

```
Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Gly Ser Met Ala
  1               5                  10                  15

Gly Val Leu Pro Ala His Gly Thr Gln His Gly Ile Arg Leu Pro Leu
             20                  25                  30

Arg Ser Gly Leu Gly Gly Ala Pro Leu Gly Leu Arg Leu Pro Arg Glu
         35                  40                  45

Thr Asp Glu Glu Pro Glu Glu Pro Gly Arg Arg Gly Ser Phe Val Glu
     50                  55                  60

Met Val Asp Asn Leu Arg Gly Lys Ser Gly Gln Gly Tyr Tyr Val Glu
 65                  70                  75                  80

Met Thr Val Gly Ser Pro Pro Gln Thr Leu Asn Ile Leu Val Asp Thr
                 85                  90                  95

Gly Ser Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu His
            100                 105                 110

Arg Tyr Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg Lys
        115                 120                 125

Gly Val Tyr Val Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu Gly
    130                 135                 140

Thr Asp Leu Val Ser Ile Pro His Gly Pro Asn Val Thr Val Arg Ala
145                 150                 155                 160

Asn Ile Ala Ala Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly Ser
                165                 170                 175

Asn Trp Glu Gly Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg Pro
            180                 185                 190

Asp Asp Ser Leu Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr His
        195                 200                 205
```

Val Pro Asn Leu Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro Leu
210 215 220

Asn Gln Ser Glu Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile Gly
225 230 235 240

Gly Ile Asp His Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro Ile
245 250 255

Arg Arg Glu Trp Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile Asn
260 265 270

Gly Gln Asp Leu Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys Ser
275 280 285

Ile Val Asp Ser Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val Phe
290 295 300

Glu Ala Ala Val Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys Phe
305 310 315 320

Pro Asp Gly Phe Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala Gly
325 330 335

Thr Thr Pro Trp Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met Gly
340 345 350

Glu Val Thr Asn Gln Ser Phe Arg Ile Thr Ile Leu Pro Gln Gln Tyr
355 360 365

Leu Arg Pro Val Glu Asp Val Ala Thr Ser Gln Asp Asp Cys Tyr Lys
370 375 380

Phe Ala Ile Ser Gln Ser Ser Thr Gly Thr Val Met Gly Ala Val Ile
385 390 395 400

Met Glu Gly Phe Tyr Val Val Phe Asp Arg Ala Arg Lys Arg Ile Gly
405 410 415

Phe Ala Val Ser Ala Cys His Val His Asp Glu Phe Arg Thr Ala Ala
420 425 430

Val Glu Gly Pro Phe Val Thr Leu Asp Met Glu Asp Cys Gly Tyr Asn
435 440 445

Ile Pro Gln Thr Asp Glu Ser Thr Leu Met Thr Ile Ala Tyr Val Met
450 455 460

```
Ala Ala Ile Cys Ala Leu Phe Met Leu Pro Leu Cys Leu Met Val Cys
465                 470                 475                 480

Gln Trp Arg Cys Leu Arg Cys Leu Arg Gln Gln His Asp Asp Phe Ala
                485                 490                 495

Asp Asp Ile Ser Leu Leu Lys
            500
```

<210> 4
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 4
Ser Glu Val Lys Met Asp Ala Glu Phe Arg
1   5   10

<210> 5
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Peptide

<400> 5
Ser Glu Val Asn Leu Asp Ala Glu Phe Arg
1   5   10

<210> 6
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Peptide

<400> 6

Ser Val Asn Met Ala Glu Gly Asp
1 5

<210> 7
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Peptide

<400> 7
Lys Gly Gly Val Val Ile Ala Thr Val Ile Val Lys
1 5 10

<210> 8
<211> 4
<212> PRT
<213> Homo sapiens

<400> 8
Asp Thr Ser Gly
1

<210> 9
<211> 8
<212> PRT
<213> Homo sapiens

<400> 9
Leu Val Asn Met Ala Glu Gly Asp
1 5

<210> 10
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 10
ggtaagcatc ccccatggcc ccaacgtc 28

<210> 11
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 11
gacgttgggg ccatggggga tgcttacc 28

<210> 12
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 12
acgttgtctt tgatcgggcc cgaaaacgaa ttgg 34

<210> 13
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 13
ccaattcgtt ttcgggcccg atcaaagaca acg 33

<210> 14
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 14
ccatcctaat acgactcact atagggc 27

<210> 15
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 15
actcactata gggctcgagc ggc 23

<210> 16
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 16
cttttgagca agttcagcct ggttaa 26

<210> 17
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 17
gaggtggctt atgagtattt cttccagggt a 31

<210> 18
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 18
tggcgacgac tcctggagcc cg 22

<210> 19
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 19
tgacaccaga ccaactggta atgg 24

<210> 20
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 20
catatggcgg gagtgctgcc tgcccac 27

<210> 21
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 21
ggatcctcac ttcagcaggg agatgtcatc agcaaagt 38

<210> 22
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Oxidized
Insulin B-chain

<220>

<223> Xaa at site 3 represents cysteic acid

<400> 22
His Leu Xaa Gly Ser His Leu Val
1 5

<210> 23
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Oxidized Insulin B-chain

<220>
<223> Xaa at site 1 represents cysteic acid

<400> 23
Xaa Gly Glu Arg Gly Phe Phe Tyr
1 5

<210> 24
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Peptide

<400> 24
Val Gly Ser Gly Val
1 5

<210> 25
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Peptide

<400> 25

Val Gly Ser Gly Val Leu Leu
1 5

<210> 26
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Peptide

<400> 26
Gly Val Leu Leu Ser Arg Lys
1 5

<210> 27
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Inhibitors

<400> 27
Val Asn Leu Ala Ala Glu Phe
1 5

<210> 28
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Inhibitors

<400> 28
Glu Val Asn Leu Ala Ala Glu Phe
1 5

<210> 29
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Peptide

<400> 29
Asn Leu Ala Ala
1

<210> 30
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Peptide

<400> 30
Val Gly Ser Gly Val Leu Leu Ser Arg Lys
1 5 10

<210> 31
<211> 326
<212> PRT
<213> Homo sapiens

<220>
<223> Amino acids 2-5, 6-9, 13-20, 25-32, 65-67, 69-74,
79-87, 89-91, 99-106, 119-122, 150-154, 164-167,
180-183, 191-194, 196-199, 201-204, 210-214,
221-223, 258-262, 265-269, and 275-278 are Beta
Strands

<220>
<223> Amino acids 281-284, 286-288, 298-301, 310-315,
and 319-324 are Beta strands

<220>
<223> Amino acids 48-51, 111-114, 136-142, 225-234,
249-254, 271-274, and 303-306 are Helices

<220>
<223> Amino acids 12-13, 30, 32, 34-35, 73-77, 111, 117,
120, 189, 213, 215, 217-220, 287, 289, 291, 298,
and 300 are residues in contact with pepstatin.

<220>
<223> Pepsin

<400> 31

```
Val Asp Glu Gln Pro Leu Glu Asn Tyr Leu Asp Met Glu Tyr Phe Gly
  1               5                  10                  15

Thr Ile Gly Ile Gly Thr Pro Ala Gln Asp Phe Thr Val Val Phe Asp
             20                  25                  30

Thr Gly Ser Ser Asn Leu Trp Val Pro Ser Val Tyr Cys Ser Ser Leu
         35                  40                  45

Ala Cys Thr Asn His Asn Arg Phe Asn Pro Glu Asp Ser Ser Thr Tyr
     50                  55                  60

Gln Ser Thr Ser Glu Thr Val Ser Ile Thr Tyr Gly Thr Gly Ser Met
 65                  70                  75                  80

Thr Gly Ile Leu Gly Tyr Asp Thr Val Gln Val Gly Gly Ile Ser Asp
                 85                  90                  95

Thr Asn Gln Ile Phe Gly Leu Ser Glu Thr Glu Pro Gly Ser Phe Leu
            100                 105                 110

Tyr Tyr Ala Pro Phe Asp Gly Ile Leu Gly Leu Ala Tyr Pro Ser Ile
        115                 120                 125

Ser Ser Ser Gly Ala Thr Pro Val Phe Asp Asn Ile Trp Asn Gln Gly
    130                 135                 140

Leu Val Ser Gln Asp Leu Phe Ser Val Tyr Leu Ser Ala Asp Asp Gln
145                 150                 155                 160

Ser Gly Ser Val Val Ile Phe Gly Gly Ile Asp Ser Ser Tyr Tyr Thr
                165                 170                 175

Gly Ser Leu Asn Trp Val Pro Val Thr Val Glu Gly Tyr Trp Gln Ile
            180                 185                 190

Thr Val Asp Ser Ile Thr Met Asn Gly Glu Ala Ile Ala Cys Ala Glu
        195                 200                 205

Gly Cys Gln Ala Ile Val Asp Thr Gly Thr Ser Leu Leu Thr Gly Pro
    210                 215                 220

Thr Ser Pro Ile Ala Asn Ile Gln Ser Asp Ile Gly Ala Ser Glu Asn
```

225 230 235 240

Ser Asp Gly Asp Met Val Val Ser Cys Ser Ala Ile Ser Ser Leu Pro
245 250 255

Asp Ile Val Phe Thr Ile Asn Gly Val Gln Tyr Pro Val Pro Pro Ser
260 265 270

Ala Tyr Ile Leu Gln Ser Glu Gly Ser Cys Ile Ser Gly Phe Gln Gly
275 280 285

Met Asn Leu Pro Thr Glu Ser Gly Glu Leu Trp Ile Leu Gly Asp Val
290 295 300

Phe Ile Arg Gln Tyr Phe Thr Val Phe Asp Arg Ala Asn Asn Gln Val
305 310 315 320

Gly Leu Ala Pro Val Ala
325

activity of presenilin therefore enhances the progression of Alzheimer's disease. This is supported by the observation that in the absence of presenilin gene, the production of alpha-beta42 peptide is lowered (De Strooper et al., Nature 391, 387 (1998)). Since unprocessed presenilin is degraded quickly, the processed, heterodimeric presenilin must be responsible for the accumulation of alpha-beta42 leading to Alzheimer's disease. The processing of presenilin by memapsin 2 would enhance the production of alpha-beta42 and therefore, further the progress of Alzheimer's disease. Therefore a memapsin 2 inhibitor that crosses the blood brain barrier can be used to decrease the likelihood of developing or slow the progression of Alzheimer's disease which is mediated by deposition of alpha-beta42. Since memapsin 2 cleaves APP at the beta cleavage site, prevention of APP cleavage at the beta cleavage site will prevent the build up of alpha-beta42.

Vaccines

The catalytically active memapsin 2 or fragments thereof including the active site defined by the presence of two catalytic aspartic residues and substrate binding cleft can be used to induce an immune response to the memapsin 2. The memapsin 2 is administered in an amount effective to elicit blocking antibodies, i.e., antibodies which prevent cleavage of the naturally occurring substrate of memapsin 2 in the brain. An unmodified vaccine may be useful in the prevention and treatment of Alzheimer's disease. The response to the vaccine may be influenced by its composition, such as inclusion of an adjuvant, viral proteins from production of the recombinant enzyme, and/or mode of administration (amount, site of administration, frequency of administration, etc). Since it is clear that the enzyme must be properly folded in order to be active, antibody should be elicited that is active against the endogenous memapsin 2. Antibodies that are effective against the endogenous enzyme are less likely to be produced against the enzyme that is not properly refolded.

Pharmaceutically Acceptable Carriers

The inhibitors will typically be administered orally or by injection. Oral administration is preferred. Alternatively, other formulations can be used for delivery by pulmonary, mucosal or transdermal routes. The inhibitor will usually be administered in combination with a pharmaceutically acceptable carrier. Pharmaceutical carriers are known to those skilled in the art. The appropriate carrier will typically be selected based on the mode of administration. Pharmaceutical compositions may also include one or more active ingredients such as antimicrobial agents, antiinflammatory agents, and analgesics.

Preparations for parenteral administration or administration by injection include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Preferred parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, and electrolyte replenishers (such as those based on Ringer's dextrose).

Formulations for topical (including application to a mucosal surface, including the mouth, pulmonary, nasal, vaginal or rectal) administration may include ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Formulations for these applications are known. For example, a number of pulmonary formulations have been developed, typically using spray drying to formulate a powder having particles with an aerodynanmic diameter of between one and three microns, consisting of drug or drug in combination with polymer and/or surfactant.

Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets, or tablets. Thickeners, flavorings, diluents, emulsifiers, dispersing aids or binders may be desirable.

Peptides as described herein can also be administered as a pharmaceutically acceptable acid- or base- addition salt, formed by reaction with inorganic acids such as hydrochloric acid, hydrobromic acid, perchloric acid, nitric acid, thiocyanic acid, sulfuric acid, and phosphoric acid, and organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, pyruvic acid, oxalic acid, malonic acid, succinic acid, maleic acid, and fumaric acid, or by reaction with an inorganic base such as sodium hydroxide, ammonium hydroxide, potassium hydroxide, and organic bases such as mono-, di-, trialkyl and aryl amines and substituted ethanolamines.

Dosages

Dosing is dependent on severity and responsiveness of the condition to be treated, but will normally be one or more doses per day, with course of treatment lasting from several days to several months or until the attending physician determines no further benefit will be obtained. Persons of ordinary skill can determine optimum dosages, dosing methodologies and repetition rates.

The dosage ranges are those large enough to produce the desired effect in which the symptoms of the memapsin 2 mediated disorder are alleviated (typically characterized by a decrease in size and/or number of amyloid plaque, or by a failure to increase in size or quantity), or in which cleavage of the alpha-beta42 peptide is decreased. The dosage can be adjusted by the individual physician in the event of any counterindications.

The present invention will be further understood by reference to the following non-limiting examples.

Example 1. Cloning of memapsin 2.

1. Cloning and nucleotide sequence of pro-memapsin 2.

New sequences homologous to human aspartic proteases were found in the following entries in the EST IMAGE database: AA136368 pregnant uterus ATCC 947471, AA207232 neurepithelium ATCC 214526, and R55398 human breast ATCC 392689. The corresponding bacterial strains: #947471, #214526, and # 392689 containing the EST sequences were obtained from the ATCC (Rockville, MD). The sequencing of these clones obtained from ATCC

confirmed that they contained sequences not identical to known human aspartic proteases. The completed sequences of these clones assembled into about 80% of prepro-M2 cDNA. Full length cDNAs of these clones were obtained using the following methods.

The Human Pancreas Marathon-Ready cDNA (Clontech), which is double-strand cDNA obtained by reverse-transcription, primer addition, and second strand synthesize of mRNA from human tissues, was used as template for PCR amplification. An adapter primer (AP1) and a nested adapter primer (AP2) were used for 5'- and 3'-RACE PCR. For PCR the 5'-region of the memapsin 2 cDNA, primers AP1 and NHASPR1 were used. Primers for the 3'-end of the cDNA are NHASPF2 and AP1. The middle of the cDNA was amplified by primers NHASPF1 and NHASPR2. The sequence for the primers is as follows: NHASPF1: GGTAAGCATCCCCCATGGCCCCAACGTC (SEQ ID NO:10),

NHASPR1: GACGTTGGGGCCATGGGGGATGCTTACC (SEQ ID NO:11),

NHASPF2: ACGTTGTCTTTGATCGGGCCCGAAAACGAATTGG (SEQ ID NO:12),

NHASPR2: CCAATTCGTTTTCGGGCCCGATCAAAGACAACG (SEQ ID NO:13),

AP1: CCATCCTAATACGACTCACTATAGGGC (SEQ ID NO:14), and

AP2: ACTCACTATAGGGCTCGAGCGGC (SEQ ID NO:15)

Memapsin 2 was also cloned from a human pancreas library (Quick-Screen Human cDNA Library Panel) contained in lambda-gt10 and lambda-gt11 vectors. The primers from the vectors, GT10FWD, GT10REV, GT11FWD, and GT11REV, were used as outside primers. The sequence of the primers used was:

GT10FWD: CTTTTGAGCAAGTTCAGCCTGGTTAA (SEQ ID NO:16),

GT10REV: GAGGTGGCTTATGAGTATTTCTTCCAGGGTA (SEQ ID NO:17),

GT11FWD: TGGCGACGACTCCTGGAGCCCG (SEQ ID NO:18),

GT11REV: TGACACCAGACCAACTGGTAATGG (SEQ ID NO:19).

In addition, memapsin 2 cDNA was amplified directly from the human pancreatic lambda-gt10 and lambda-gt11 libraries. The sequence of the primers was: PASPN1: catatgGCGGGAGTGCTGCCTGCCCAC (SEQ ID NO:20) and

NHASPC1: ggatccTCACTTCAGCAGGGAGATGTCATCAGCAAAGT (SEQ ID NO:21).

The amplified memapsin 2 fragments were cloned into an intermediate PCR vector (Invitrogen) and sequenced.

The assembled cDNA from the fragments, the nucleotide and the deduced protein sequence are shown in SEQ ID NO 1 and SEQ ID NO 2.

Pro-memapsin 2 is homologous to other human aspartic.proteases. Based on the alignments, Pro-memapsin 2 contains a *pro* region, an aspartic protease region, and a trans-membrane region near the C-terminus. The active enzyme is memapsin 2 and its pro-enzyme is pro-memapsin 2.

Example 2. Distribution of memapsin 2 in human tissues.

Multiple tissue cDNA panels from Clontech were used as templates for PCR amplification of a 0.82 kb fragment of memapsin 2 cDNA. The primers used for memapsin 2 were NHASPF1 and NHASPR2. Tissues that contain memapsin 2 or fragments of memapsin 2 yielded amplified PCR products. The amount of amplified product indicated that memapsin 2 is present in the following organs from most abundant to least abundant: pancreas, brain, lung, kidney, liver, placenta, and heart. Memapsin 2 is also present in spleen, prostate, testis, ovary, small intestine, and colon cells.

Example 3. Expression of pro-memapsin 2 cDNA in E. coli, refolding and purification of pro-memapsin 2.

The pro-memapsin 2 was PCR amplified and cloned into the *Bam*HI site of a pET11a vector. The resulting vector expresses pro-memapsin 2 having a sequence from Ala-8p to Ala 326. Figure 1 shows the construction of two expression vectors, pET11-memapsin 2-T1 (hereafter T1) and pET11-memapsin 2-T2 (hereafter T2). In both vectors, the N-terminal 15 residues of the expressed recombinant proteins are derived from the expression vector. Pro-

memapsin 2 residues start at residue Ala-16. The two recombinant pro-memapsin 2s have different C-terminal lengths. Clone T1 ends at Thr- 454 and clone T2 ends at Ala-419. The T1 construct contains a C-terminal extension from the T2 construct but does not express any of the predicted transmembrane domain.

Expression of recombinant proteins and recovery of inclusion bodies

The T1 and T2 expression vectors were separately transfected into *E. coli* strain BL21(DE3). The procedures for the culture of transfected bacteria, induction for synthesis of recombinant proteins and the recovery and washing of inclusion bodies containing recombinant proteins are essentially as previously described (Lin et al., 1994).

Three different refolding methods have produced satisfactory results.

(i) The rapid dilution method.

Pro-memapsin 2 in 8 M urea/100 mM beta-mercaptoethanol with OD_{280nm} = 5 was rapidly diluted into 20 volumes of 20 mM-Tris, pH 9.0. The solution was slowly adjusted into pH 8 with 1 M HCl. The refolding solution was then kept at 4° C for 24 to 48 hours before proceeding with purification.

(ii) The reverse dialysis method

An equal volume of 20 mM Tris, 0.5 mM oxidized/1.25 mM reduced glutathione, pH 9.0 is added to rapidly stirred pro-memapsin 2 in 8 M urea/10 mM beta-mercaptoethanol with $OD_{280\ nm}$ = 5. The process is repeated three more times with 1 hour intervals. The resulting solution is then dialyzed against sufficient volume of 20 mM Tris base so that the final urea concentration is 0.4 M. The pH of the solution is then slowly adjusted to 8.0 with 1 M HCl.

iii. The preferred method for refolding.

Inclusion bodies are dissolved in 8 M urea, 0.1 M Tris, 1 mM Glycine, 1 mM EDTA, 100 mM beta-mercaptoethanol, pH 10.0. The OD_{280} of the inclusion bodies are adjusted to 5.0 with the 8 M urea solution without beta-mercaptoethanol. The final solution contains the following reducing reagents: 10 mM beta-mercaptoethanol, 10 mM DTT (Dithiothreitol), 1 mM reduced glutathion, and 0.1 M oxidized glutathion. The final pH of the solution is 10.0.

The above solution is rapidly diluted into 20 volumes of 20 mM Tris base, the pH is adjusted to 9.0, and the resulting solution is kept at 4 °C for 16 hr. The solution is equilibrated to room temperature in 6 hr, and the pH is adjusted to 8.5. The solution is returned to 4 °C again for 18 hr.

The solution is again equilibrated to room temperature in 6 hr, and the pH is adjusted to 8.0. The solution is returned to 4 °C again for 4 to 7 days.

The refolding procedures are critical to obtain an enzymically active preparation which can be used for studies of subsite specificity of M2, to analyze inhibition potency of M2 inhibitors, to screen for inhibitors using either random structural libraries or existing collections of compound libraries, to produce crystals for crystallography studies of M2 structures, and to produce monoclonal or polyclonal antibodies of M2.

Purification of recombinant pro-memapsin 2-T2

The refolded material is concentrated by ultrafiltration, and separated on a SEPHACRYL™ S-300 column equilibrated with 20 mM Tris.HCl, 0.4 M urea, pH 8.0. The refolded peak (second peak) from the S-300 column can be further purified with a FPLC RESOURCE-Q™ column, which is equilibrated with 20 mM Tris-HCl, 0.4 M urea, pH 8.0. The enzyme is eluted from the column with a linear gradient of NaCl. The refolded peak from S-300 can also be activated before further purification. For activation, the fractions are mixed with equal volume 0.2 M Sodium Acetate, 70% glycerol, pH 4.0. The mixture is incubated at 22 °C for 18 hr, and then dialyzed twice against 20 volumes of 20 mM Bis-Tris, 0.4 M urea, pH 6.0. The dialyzed materials are then further purified on a FPLC RESOURCE-Q™ column equilibrated with 20 Bis-Tris, 0.4 M urea, pH 6.0. The enzyme is eluted with a linear gradient of NaCl.

SDS-PAGE analysis of the S-300 fractions under reduced and non-reduced conditions indicated that Pro-memapsin 2 first elutes as a very high molecular weight band (greater than about 42 kD) under non-reduced conditions. This indicates that the protein is not folded properly in these fractions, due to disulfide cross linking of proteins. Subsequent fractions contain a protein of predicted pro-memapsin 2-T2 size (about 42 kDa). The pro-

enzyme obtained in these fractions is also proteolytically active for auto-catalyzed activation. These fractions were pooled and subjected to chromatography on the FPLC RESOURCE™ column eluted with a linear gradient of NaCl. Some fractions were analyzed using SDS-PAGE under non-reducing conditions. The analysis showed that fractions 6 and 7 contained most of the active proteins, which was consistent with the first FPLC peak containing the active protein. The main peak was coupled to a shoulder peak, and was present with repeated purification with the same RESOURCE™ Q column. The main shoulder peaks were identified as active pro-memapsin 2 that exist in different conformations under these conditions.

Example 4. Proteolytic activity and cleavage-site preferences of recombinant memapsin 2.

The amino acid sequence around the proteolytic cleavage sites was determined in order to establish the specificity of memapsin 2. Recombinant pro-memapsin 2-T1 was incubated in 0.1 M sodium acetate, pH 4.0, for 16 hours at room temperature in order to create autocatalyzed cleavages. The products were analyzed using SDS-polyacrylamide gel electrophoresis. Several bands which corresponded to molecular weights smaller than that of pro-memapsin 2 were observed. The electrophoretic bands were trans-blotted onto a PVDF membrane. Four bands were chosen and subjected to N-terminal sequence determination in a Protein Sequencer. The N-terminal sequence of these bands established the positions of proteolytic cleavage sites on pro-memapsin 2.

In addition, the oxidized ß-chain of bovine insulin and two different synthetic peptides were used as substrates for memapsin 2 to determine the extent of other hydrolysis sites. These reactions were carried out by auto-activated pro-memapsin 2 in 0.1 M sodium acetate, pH 4.0, which was then incubated with the peptides. The hydrolytic products were subjected to HPLC on a reversed phase C-18 column and the eluent peaks were subjected to electrospray mass spectrometry for the determination of the molecular weight of the fragments. Two hydrolytic sites were identified on oxidized insulin B-chain

(Table1). Three hydrolytic sites were identified from peptide NCH-gamma. A single cleavage site was observed in synthetic peptide PS1-gamma, whose sequence (LVNMAEGD) (SEQ ID NO:9) is derived from the beta-processing site of human presenilin 1 (Table 1).

Table 1: Substrate Specificity of Memapsin 2

Site #	Substrate	P4	P3	P2	P1	P1'	P2'	P3'	P4'	
1	Pro-memapsin 2	R	G	S	M	A	G	V	L	aa 12-18 of SEQ ID No.3
2		G	T	Q	H	G	I	R	L	aa 23-30 of SEQ ID No. 3
3		S	S	N	F	A	V	G	A	aa 98-105 of SEQ ID No. 3
4		G	L	A	Y	A	E	I	A	aa 183-190 of SEQ ID No.3
5	Oxidized insulin B-chain '	H	L	C^	G	S	H	L	V	C^ is cysteic acid; SEQ ID No. 22
6		C^	G	E	R	G	F	F	Y	SEQ ID No. 23
7	Synthetic peptide*				V	G	S	G	V	Three sites cleaved in a peptide: VGSGVLLSRK (SEQ ID NO:30)
8			V	G	S	G	V	L	L	SEQ ID No. 24
9		G	V	L	L	S	R	K		SEQ ID No. 25 SEQ ID No. 26
10	Peptide**	L	V	N	M	A	E	G	D	SEQ ID No. 9

Example 5. Activation of pro-memapsin 2 and enzyme kinetics.

Incubation in 0.1 M sodium acetate, pH 4.0, for 16 h at 22°C auto-catalytically converted *pro*-$M2_{pd}$ to $M2_{pd}$. For initial hydrolysis tests, two synthetic peptides were separately incubated with *pro*-$M2_{pd}$ in 0.1 M Na acetate, pH 4.0 for different periods ranging from 2 to 18 h. The incubated samples

were subjected to LC/MS for the identification of the hydrolytic products. For kinetic studies, the identified HPLC (Beckman System Gold) product peaks were integrated for quantitation. The K_m and k_{cat} values for presenilin 1 and Swedish APP peptides (Table 1) were measured by steady-state kinetics. The individual K_m and k_{cat} values for APP peptide could not be measured accurately by standard methods, so its k_{cat}/K_m value was measured by competitive hydrolysis of mixed substrates against presenilin 1 peptide (Fersht, A. "Enzyme Structure and Mechanism", 2nd Ed., W.H. Freeman and Company, New York. (1985)).

The results are shown in Figures 2A and 2B. The conversion of *pro*-$M2_{pd}$ at pH 4.0 to smaller fragments was shown by SDS-polyacrylamide electrophoresis. The difference in migration between *pro*-$M2_{pd}$ and converted enzyme is evident in a mixture of the two. Figure 2A is a graph of the initial rate of hydrolysis of synthetic peptide swAPP (see Table 1) by $M2_{pd}$ at different pH. Figure 2B is a graph of the relative k_{cat}/K_m values for steady-state kinetic of hydrolysis of peptide substrates by $M2_{pd}$.

Example 6. Expression in Mammalian cells.

Methods

PM2 cDNA was cloned into the *Eco*RV site of vector pSecTag A (Invitrogen). Human APP cDNA was PCR amplified from human placenta 8-gt11 library (Clontech) and cloned into the *Nhe*I and *Xba*I sites of pSecTag A. The procedure for transfection into HeLa cells and vaccinia virus infection for T7-based expression are essentially the same as described by Lin, X., *FASEB J.* 7:1070-1080 (1993).

Transfected cells were metabolically labeled with 200 microCi ^{35}S methionine and cysteine (TransLabel; ICN) in 0.5 ml of serum-free/methionine-free media for 30 min, rinsed with 1 ml media, and replaced with 2 ml DMEM/10% FCS. In order to block vesicle acidification, Bafilomycin A1 was included in the media (Perez, R.G., et al., *J Biol. Chem* 271:9100-9107 (1996)). At different time points (chase), media was removed and the cells were harvested and lysed in 50 mM Tris, 0.3 M NaCl, 5 mM EDTA, 1% Triton X-

100, pH 7.4, containing 10 mM iodoacetamide, 10 :M TPCK, 10 :M TLCK, and 2 microg/ml leupeptin. The supernatant (14,000 x g) of cell lysates and media were immunoadsorbed onto antibody bound to protein G sepharose (Sigma). Anti-APP N-terminal domain antibody (Chemicon) was used to recover the betaN-fragment of APP and anti-alpha-beta $_{1-17}$ antibody (Chemicon, recognizing the N-terminal 17 residues of alpha-beta) was used to recover the 12 kDa ß C-fragment. The former antibody recognized only denatured protein, so media was first incubated in 2 mM dithiothrietol 0.1% SDS at 55°C for 30 min before immunoabsorption. Samples were cooled and diluted with an equal volume of cell lysis buffer before addition of anti-APP N-terminal domain (Chemicon). Beads were washed, eluted with loading buffer, subjected to SDS-PAGE (NOVEX™) and visualized by autoradiogram or phosphorimaging (Molecular Dynamics) on gels enhanced with Amplify (Amersham). Immunodetection of the betaN-fragment was accomplished by transblotting onto a PVDF membrane and detecting with anti-alpha-beta$_{1-17}$ and chemiluminescent substrate (Amersham).

Results.

HeLa cells transfected with APP or M2 in 4-well chamber slides were fixed with acetone for 10 min and permeabilized in 0.2% Triton X-100 in PBS for 6 min. For localizing M2, polyclonal goat anti-*pro*-M2$_{pd}$ antibodies were purified on DEAE-sepharose 6B and affinity purified against recombinant *pro*-M2$_{pd}$ immobilized on Affigel (BioRad). Purified anti-*pro*-M2$_{pd}$ antibodies were conjugated to Alexa568 (Molecular Probes) according to the manufacturer's protocol. Fixed cells were incubated overnight with a 1:100 dilution of antibody in PBS containing 0.1% BSA and washed 4 times with PBS. For APP, two antibodies were used. Antibody A ß $_{1-17}$ (described above) and antibody Aß $_{17-42}$, which recognizes the first 26 residues following the beta-secretase cleavage site (Chemicon). After 4 PBS washes, the cells were incubated overnight with an anti-mouse FITC conjugate at a dilution of 1:200. Cells were mounted in Prolong anti-fade reagent (Molecular Probes) and visualized on a Leica TCS confocal laser scanning microscope.

Example 7: Design and Synthesis of OM99-1 and OM99-2.

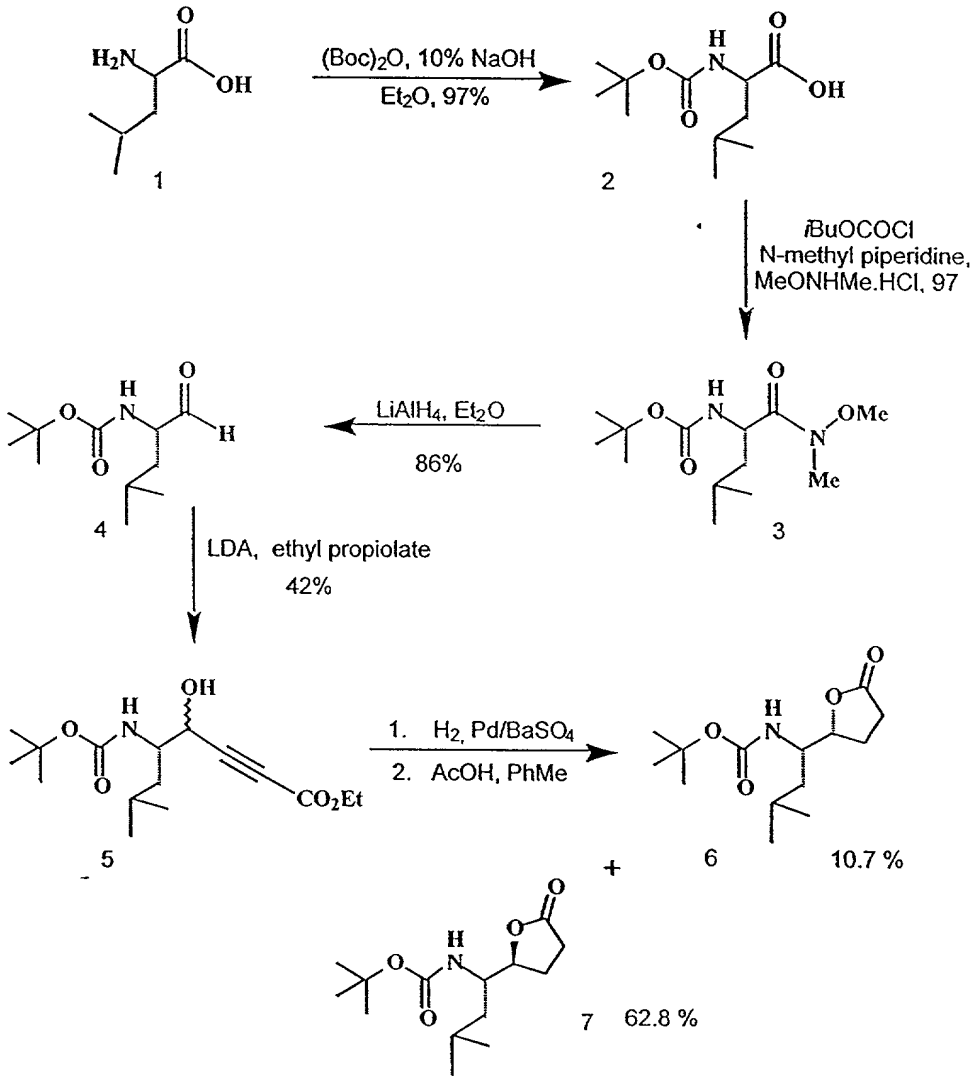
Based on the results of specificity studies of memapsin 2, it was predicted that good residues for positions P1 and P1' would be Leu and Ala. It was subsequently determined from the specificity data that P1' preferred small residues, such as Ala and Ser. However, the crystal structure (determined below in Example 9) indicates that this site can accommodate a lot of larger residues. It was demonstrated that P1' of memapsin 2 is the position with the most stringent specificity requirement where residues of small side chains, such as Ala, Ser, and Asp, are preferred. Ala was selected for P1' mainly because its hydophobicity over Ser and Asp is favored for the penetration of the blood-brain barrier, a requirement for the design of a memapsin 2 inhibitor drug for treating Alzheimer's disease. Therefore, inhibitors were designed to place a transition-state analogue isostere between Leu and Ala (shown as Leu*Ala, where * represents the transition-state isostere, -CH(OH)-CH_2-) and the subsite P4, P3, P2, P2', P3' and P4' are filled with the beta-secretase site sequence of the Swedish mutant from the beta-amyloid protein. The structures of inhibitors OM99-1 and OM99-2 are shown below and in Figures 3A and 3B, respectively:

OM99-1: Val-Asn-Leu*Ala-Ala-Glu-Phe (SEQ. ID NO. 27)

OM99-2: Glu-Val-Asn-Leu*Ala-Ala-Glu-Phe (SEQ. ID NO. 28)

The Leu*Ala dipeptide isostere was synthesized as follows:

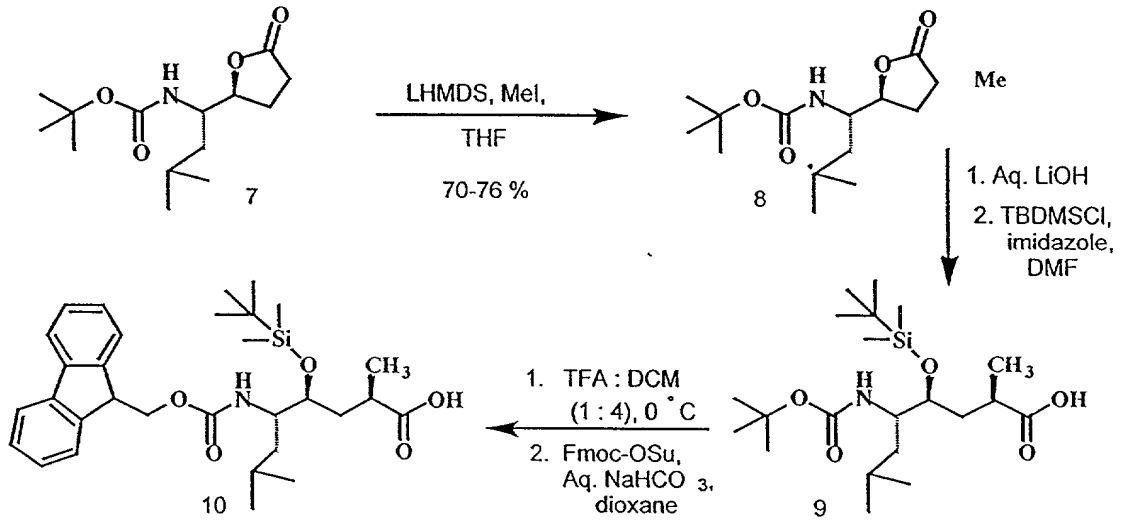
The Leu-Ala dipeptide isostere for the M_2-inhibitor was prepared from L-leucine. As shown in Scheme 1, L-leucine was protected as its BOC-derivative 2 by treatment with BOC_2O in the presence of 10% NaOH in diethyl ether for 12 h. Boc-leucine 2 was then converted to Weinreb amide 3 by treatment with isobutyl chcloroformate and N-methylpiperidine followed by treatment of the resulting mixed anhydride with N,O-dimethylhydroxylamine



(Scheme 1)

(Nahm and Weinreb, Tetrahedron Letters 1981, 32, 3815). Reduction of 3 with lithium aluminum hydride in diethyl ether provided the aldehyde 4. Reaction of the aldehyde 4 with lithium propiolate derived from the treatment of ethyl

propiolate and lithium diisopropylamide afforded the acetylenic alcohol 5 as an inseparable mixture of diastereomers (5.8:1) in 42% isolated yield (Fray, Kaye and Kleinman, J. Org. Chem. 1986, 51, 4828-33). Catalytic hydrogenation of 5 over $Pd/BaSO_4$ followed by acid-catalyzed lactonization of the resulting gamma-hydroxy ester with a catalytic amount of acetic acid in toluene at reflux, furnished the gamma-lactone 6 and 7 in 73% yield. The isomers were separated by silica gel chromatography by using 40% ethyl acetate in hexane as the eluent.



(Scheme 2)

Introduction of the methyl group at C-2 was accomplished by stereoselective alkylation of 7 with methyl iodide (Scheme 2). Thus, generation of the dianion of lactone 7 with lithium hexamethyldisilazide (2.2 equivalents) in tetrahydrofuran at -78°C (30 min) and alkylation with methyl iodide (1.1 equivalents) for 30 min at -78°C, followed by quenching with propionic acid (5 equivalents), provided the desired alkylated lactone 8 (76% yield) along with a small amount (less than 5%) of the corresponding epimer (Ghosh and Fidanze, 1998 J. Org. Chem. 1998, 63, 6146-54). The epimeric cis-lactone was removed by column chromatography over silica gel using a mixture (3:1) of ethyl acetate and hexane as the solvent system. The stereochemical assignment of alkylated lactone 8 was made based on extensive ^{1}H-NMR NOE experiments. Aqueous lithium hydroxide promoted hydrolysis of the lactone 8 followed by protection of the gamma-hydroxyl group with *tert*-butyldimethylsilyl chloride in the

presence of imidazole and dimethylaminopyridine in dimethylformamide afforded the acid 9 in 90% yield after standard work-up and chromatography. Selective removal of the BOC-group was effected by treatment with trifluoroacetic acid in dichloromethane at 0°C for 1 h. The resulting amine salt was then reacted with commercial (Aldrich, Milwaukee) Fmoc-succinimide derivative in dioxane in the presence of aqueous $NaHCO_3$ to provide the Fmoc-protected L*A isostere 10 in 65% yield after chromatography. Protected isostere 10 was utilized in the preparation of a random sequence inhibitor library.

Experimental procedure

N-(tert-Butoxycarbonyl)-L-Leucine (2).

To the suspension of 10 g (76.2 mmol) of L-leucine in 140 mL of diethyl ether was added 80 mL of 10 % NaOH. After all solid dissolves, 20 mL (87.1 mmol) of BOC_2O was added to the reaction mixture. The resulting reaction mixture was stirred at 23°C for 12 h. After this period, the layers were separated and the aqueous layer was acidified to pH 1 by careful addition of 1 N aqueous HCl at 0 °C. The resulting mixture was extracted with ethyl acetate (3 x 100 mL). The organic layers were combined and washed with brine and dried over anhydrous Na_2SO_4. The solvent was removed under reduced pressure to provide title product which was used directly for next reaction without further purification (yield, 97 %). ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.89 (broad d, 1H, J = 8.3 Hz), 4.31 (m, 1H), 1.74-1.49 (m, 3H), 1.44 (s, 9H), 0.95 (d, 6H, J = 6.5 Hz). -

N-(tert-Butoxycarbonyl)-L-leucine-N'-methoxy-N'-methyla-mide (3).

To a stirred solution of N,O-dimethylhydroxyamine hydrochloride (5.52 g, 56.6 mmol) in dry dichloromethane (25 mL) under N_2 atmosphere at 0°C, -methylpiperidine (6.9 mL, 56.6 mmol) was added dropwise. The resulting mixture was stirred at 0°C for 30 min. In a separate flask, N-(tert-butyloxycarbonyl)-L-leucine (1) (11.9 g, 51.4 mmol) was dissolved in a mixture of THF (45 mL) and dichloromethane (180 mL) under N_2 atmosphere. The resulting solution was cooled to –20°C. To this solution was added 1-

methylpiperidine (6.9 mL, 56.6 mmol) followed by isobutyl chloroformate (7.3 mL, 56.6 mmol). The resulting mixture was stirred for 5 minutes at –20°C and the above solution of N,O-dimethylhydroxyamine was added to it. The reaction mixure was kept -20 °C for 30 minutes and then warmed to 23°C. The reaction was quenched with water and the layers were seperated. The aqueous layer was extracted with dichloromethane (3 x 100 mL). The combined organic layers were washed with 10% citric acid, saturated sodium bicarbonate, and brine. The organic layer was dried over anhydrous Na_2SO_4 and concentrated under the reduced pressure. The residue was purified by flash silica gel chromatography (25% ethyl acetate/hexane) to yield the title compound 3 (13.8 g, 97%) as a pale yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.06 (broad d, 1H, J = 9.1 Hz), 4.70 (m, 1H), 3.82 (s, 3H), 3.13 (s, 3H), 1.70 (m, 1H), 1.46-1.36 (m, 2H) 1.41 (s, 9H), 0.93 (dd, 6H, J = 6.5, 14.2 Hz).

N-(tert-Butoxycarbonyl)-L-leucinal (4).

To a stirred suspension of lithium aluminum hydride (770 mg, 20.3 mmol) in dry diethyl ether (60 mL) at -40 °C under N_2 atmosphere, was added N-tert-butyloxycarbonyl-L-leucine-N'-methoxy-N'-methylamide (5.05 g, 18.4 mmol) in diethyl ether (20 mL). The resulting reaction mixture was stirred for 30 min. After this period, the reaction was quenched with 10% $NaHSO_4$ solution (30 mL). The resulting reaction mixture was then warmed to 23°C and stirred at that temperature for 30 min. The resulting solution was filtered and the filter cake was washed by two portions of diethyl ether. The combined organic layers were washed with saturated sodium bicarbonate, brine and dried over anhydrous $MgSO_4$. Evaporation of the solvent under reduced pressure afforded the title aldehyde 4 (3.41 g) as a pale yellow oil. The resulting aldehyde was used immediately without further purification. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.5 (s, 1H), 4.9 (s, 1H), 4.2 (broad m, 1H), 1.8-1.6 (m, 2H), 1.44 (s, 9H), 1.49-1.39 (m, 1H), 0.96 (dd, 6H, J = 2.7, 6.5 Hz).

Ethyl (4S,5S)-and (4R,5S)-5-[(tert-Butoxycarbonyl)amino]-4-hydroxy-7-methyloct-2-ynoate (5).

To a stirred solution of diisopropylamine (1.1 mL, 7.9 mmol) in dry THF (60 mL) at 0°C under N_2 atmosphere, was added n-BuLi (1.6 M in hexane, 4.95 mL, 7.9 mmol) dropwise. The resulting solution was stirred at 0°C for 5 min and then warmed to 23°C and stirred for 15 min. The mixture was cooled to -78°C and ethyl propiolate (801 μL) in THF (2 mL) was added dropwise over a period of 5 min. The mixture was stirred for 30 min, after which N-Boc-L-leucinal 4 (1.55 g, 7.2 mmol) in 8 mL of dry THF was added. The resulting mixture was stirred at -78°C for 1 h. After this period, the reaction was quenched with acetic acid (5 mL) in THF (20 mL). The reaction mixure was warmed up to 23°C and brine solution was added. The layers were separated and the organic layer was washed with saturated sodium bicarbonate and dried over Na_2SO_4. Evaporation of the solvent under reduced pressure provided a residue which was purified by flash silica gel chromatography (15 % ethyl acetate / hexane) to afford a mixture (3:1) of acetylenic alcohols 5 (0.96 g, 42 %). ^{1}H NMR (300 MHz, $CDCl_3$) δ 4.64 (d, 1H, J = 9.0 Hz), 4.44 (broad s, 1H), 4.18 (m, 2H), 3.76 (m, 1H), 1.63 (m, 1H), 1.43-1.31 (m, 2H), 1.39 (s, 9H), 1.29-1.18 (m, 3H), 0.89 (m, 6H).

(5S,1'S)-5-[1'-[(tert-Butoxycarbonyl)amino]-3'-methylbutyl] -dihydrofuran-2(3H)-one (7).

To a stirred solution of the above mixture of acetylenic alcohols (1.73 g, 5.5 mmol) in ethyl acetate (20 mL) was added 5% Pd/$BaSO_4$ (1 g). The resulting mixture was hydrogenated at 50 psi for 1.5 h. After this period, the catalyst was filtered off through a plug of Celite and the filtrate was concentrated under reduced pressure. The residue was dissolved in toluene (20 mL) and acetic acid (100 μL). The reaction mixure was refluxed for 6 h. After this period, the reaction was cooled to 23°C and the solvent was evaporated to give a residue which was purified by flash silica gel chromatography (40% diethyl ether / hexane) to yield the (5S, 1S')-gamma-lactone 7 (0.94 g, 62.8 and the (5R, 1S')-gamma-lactone 6 (0.16 g, 10.7 %). Lactone 7: ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.50-4.44 (m, 2H), 3.84-3.82 (m, 1H), 2.50 (t, 2H, J = 7.8 Hz), 2.22-2.10 (m, 2H), 1.64-1.31 (m, 3H), 1.41 (s, 9H), 0.91 (dd, 6H, J = 2.2, 6.7

Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ 177.2, 156.0, 82.5, 79.8, 51.0, 42.2, 28,6, 28.2, 24.7, 24.2, 23.0, 21.9.

(3R,5S,1'S)-5-[1'-[(tert-Butoxycarbonyl)amino)]-3'-methylbut-yl]-3-methyl dihydrofuran-2(3H)-one (8).

To a stirred solution of the lactone 7 (451.8 mg, 1.67 mmol) in dry THF (8 mL) at -78°C under N_2 atmosphere, was added lithium hexamethyldisilazane (3.67 mL, 1.0 M in THF) over a period of 3 min. The resulting mixture was stirred at -78°C for 30 min to generate the lithium enolate. After this period, MeI (228 μL) was added dropwise and the resulting mixture was stirred at – 78°C for 20 min. The reaction was quenched with saturated aqueous NH_4Cl solution and was allowed to warm to 23°C. The reaction mixture was concentrated under reduced pressure and the residue was extracted with ethyl acetate (3 x 100 mL). The combined organic layers were washed with brine and dried over anhydrous Na_2SO_4. Evaporation of the solvent afforded a residue which was purified by silica gel chromatography (15 % ethyl acetate / hexane) to furnish the alkylated lactone 8 (0.36 g, 76 %) as an amorphous solid. ^1H NMR (300 MHz, $CDCl_3$) δ 4.43 (broad t, 1H, J = 6.3 Hz), 4.33 (d, 1H, J = 9.6 Hz), 3.78 (m, 1H), 2.62 (m, 1H), 2.35 (m, 1H), 1.86 (m , 1H), 1.63-1.24 (m, 3H), 1.37 (s, 9H), 1.21 (d, 3H, J = 7.5 Hz), 0.87 (dd, 6H, J = 2.6, 6.7 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ 180.4, 156.0, 80.3, 79.8, 51.6, 41.9, 34.3, 32.5, 28.3, 24.7, 23.0, 21.8, 16.6.

(2R,4S,5S)-5-[(tert-Butoxycarbonyl)amino]-4-[(tert-butyldimeth -ylsilyl)oxy]-2,7-dimethyloctanoic acid (9).

To a stirred solution of lactone 8 (0.33 g, 1.17 mmol) in THF (2 mL) was added 1 N aqueous LiOH solution (5.8 mL). The resulting mixture was stirred at 23°C for 10 h. After this period, the reaction mixture was concentrated under reduced pressure and the remaining aqueous residue was cooled to 0°C and acidified with 25% citric acid solution to pH 4. The resulting acidic solution was extracted with ethyl acetate (3 x 50 mL). The combined organic layers were washed with brine, dried over Na_2SO_4 and concentrated to yield the

corresponding hydroxy acid (330 mg) as a white foam. This hydroxy acid was used directly for the next reaction without further purification.

To the above hydroxy acid (330 mg, 1.1 mmol) in anhydrous DMF was added imidazole (1.59 g, 23.34 mmol) and tert-butyldimethylchlorosilane (1.76 g, 11.67 mmol). The resulting mixture was stirred at 23°C for 24 h. After this period, MeOH (4 mL) was added and the mixture was stirred for 1 h. The mixure was diluted with 25% citric acid (20 mL) and was extracted with ethyl acetate (3 x 20 mL). The combined extracts were washed with water, brine and dried over anhydrous Na_2SO_4. Evaporation of the solvent gave a viscous oil which was purified by flash chromatography over silica gel (35% ethyl acetate / hexane) to afford the silyl protected acid 9 (0.44 g, 90 %). IR (neat) 3300-3000 (broad), 2955, 2932, 2859, 1711 cm^{-1}; ^1H NMR (400 MHz, DMSO-d^6, 343 K) delta 6.20 (broad s, 1 H), 3.68 (m, 1H), 3.51 (broad s, 1H), 2.49-2.42 (m, 1H), 1.83 (t, 1H, J = 10.1 Hz), 1.56 (m, 1H), 1.37 (s, 9H), 1.28-1.12 (m, 3H), 1.08 (d, 3H, J = 7.1 Hz), 0.87 (d, 3H, J = 6.1 Hz) 0.86 (s, 9 H), 0.82 (d, 3H, J = 6.5 Hz), 0.084 (s, 3H), 0.052 (s, 3H).

(2R,4S,5S)-5-[(fluorenylmethyloxycarbonyl)amino]-4-[(tert-butyldi-methyl silyl)oxy]-2,7-dimethyloctanoic acid (10).

To a stirred solution of the acid 9 (0.17 g , 0.41 mmol) in dichloromethane (2 mL) at 0°C was added trifluoroacetic acid (500 µL). The resulting mixture was stirred at 0°C for 1 h and an additional portion (500 µL) of trifluoroacetic acid was added to the reaction mixture. The mixture was stirred for an additional 30 min and the progress of the reaction was monitored by TLC. After this period, the solvents were carefully removed under reduced pressure at a bath temperature not exceeding 5°C. The residue was dissolved in dioxane (3 mL) and $NaHCO_3$ (300 mg) in 5 mL of H_2O. To this solution was added Fmoc-succinimide (166.5 mg, 0.49 mmol) in 5 mL of dioxane. The resulting mixture was stirred at 23°C for 8 h. The mixure was then diluted with H_2O (5 mL) and acidified with 25% aqueous citric acid to pH 4. The acidic solution was extracted with ethyl acetate (3 x 50 mL). The combined extracts were washed with brine, dried over Na_2SO_4 and concentrated under reduced

pressure to give a viscous oil residue. Purification of the residue by flash chromatography over silica gel afforded the Fmoc-protected acid 10 (137 mg, 61%) as a white foam. ^{1}H NMR (400 MHz, DMSO-d^6, 343 K) δ 7.84 (d, 2H, J = 7.4 Hz), 7.66 (d, 2H, J = 8 Hz), 7.39 (t, 2H, J = 7.4 Hz), 7.29 (m, 2H), 6.8 (s, 1H), 4.29-4.19 (m, 3H), 3.74-3.59 (m, 2H), 2.49 (m, 1H), 1.88 (m, 1H), 1.58 (m, 1H), 1.31-1.17 (m, 3H), 1.10 (d, 3H, J = 7.1 Hz), 0.88 (s, 9H), 0.82 (d, 6H, J = 6.2 Hz), 0.089 (s, 3 H), 0.057 (s, 3H).

The synthesis of OM99-1 and OM99-2 were accomplished using solid state peptide synthesis procedure in which Leu*Ala was incorporated in the fourth step. The synthesized inhibitors were purified by reverse phase HPLC and their structure confirmed by mass spectrometry.

Example 8. Inhibition of Memapsin 2 by OM99-1 and OM99-2.

Enzyme activity was measured as described above, but with the addition of either OM99-1 or OM99-2. OM99-1 inhibited recombinant memapsin 2 as shown in Figure 5A. The Ki calculated is 3 x 10^{-8} M. The substrate used was a synthetic fluorogenic peptide substrate. The inhibition of OM99-2 on recombinant memapsin 2 was measured using the same fluorogenic substrate. The Ki value was determined to be 9.58 x 10^{-9} M, as shown in Figure 5B.

These results demonstrate that the predicted subsite specificity is accurate and that inhibitors can be designed based on the predicted specificity.

The residues in P1 and P1' are very important since the M2 inhibitor must penetrate the blood-brain barrier (BBB). The choice of Ala in P1' facilitates the penetration of BBB. Analogues of Ala side chains will also work. For example, in addition to the methyl side chain of Ala, substituted methyl groups and groups about the same size like methyl or ethyl groups can be substituted for the Ala side chain. Leu at P1 can also be substituted by groups of similar sizes or with substitutions on Leu side chain. For penetrating the BBB, it is desirable to make the inhibitors smaller. One can therefore use OM99-1 as a starting point and discard the outside subsites P4, P3, P3' and P4' . The retained structure Asn-Leu*Ala-Ala (SEQ ID NO:29) is then further evolved with substitutions for a tight-binding M2 inhibitor which can also penetrate the BBB.

Example 9. Crystallization and X-ray diffraction study of the protease domain of human memapsin 2 complexed to a specifically designed inhibitor, OM99-2.

The crystallization condition and preliminary x-ray diffraction data on recombinant human memapsin 2 complexed to OM99-2 were determined.

Production of Recombinant Memapsin 2

About 50 mg of recombinant memapsin 2 was purified as described in Example 3. For optimal crystal growth, memapsin 2 must be highly purified. Memapsin 2 was over-expressed from vector pET11a-M2pd. This memapsin 2 is the zymogen domain which includes the pro and catalytic domains to the end of the C-terminal extension but does not include the transmembrane and the intracellular domains. The vector was transfected into E. coli BL21 (DE3) and plated onto ZB agar containing 50 mg/liter ampicillin. A single colony was picked to inoculate 100 ml of liquid ZB containing 5 mg ampicillin and cultured at 30 °C, for 18 hours, with shaking at 220 RPM. Aliquots of approximately 15 ml of the overnight culture were used to inoculate each 1 liter of LB containing 50 mg of ampicillin. Cultures were grown at 37 °C, with shaking at 180 RPM, until an optical density at 600 nm near 0.8 was attained. At that time, expression was induced by addition of 119 mg of IPTG to each liter of culture. Incubation was continued for 3 additional hours post-induction.

Bacteria were harvested, suspended in 50 mM Tris, 150 mM NaCl, pH 7.5 (TN buffer), and lysed by incubation with 6 mg lysozyme for 30 minutes, followed by freezing for 18 hours at -20 °C. Lysate was thawed and made to 1 mM $MgCl_2$ then 1000 Kunitz units of DNAse were added with stirring, and incubated for 30 min. Volume was expanded to 500 ml with TN containing 0.1 % Triton X-100 (TNT buffer) and lysate stirred for 30 minutes. Insoluble inclusion bodies containing greater than 90% memapsin 2 protein were pelleted by centrifugation, and washed by resuspension in TNT with stirring for 1-2 hours. Following three additional TNT washes, the memapsin 2 inclusion bodies were dissolved in 40 ml of 8 M urea, 1 mM EDTA, 1 mM glycine, 100 mM Tris base, 100 mM beta-mercaptoethanlol (8 M urea buffer). Optical

density at 280 nm was measured, and volume expanded with 8 M urea buffer to achieve final O.D. near 0.5, with addition of sufficient quantity of beta-mercaptoethanol to attain 10 mM total, and 10 mM DTT, 1 mM reduced glutathione, 0.1 mM oxidized glutathione. The pH of the solution was adjusted to 10.0 or greater, and divided into four aliquots of 200 ml each. Each 200 ml was rapidly-diluted into 4 liters of 20 mM Tris base, with rapid stirring. The pH was adjusted immediately to 9.0, with 1 M HCl, and stored at 4 °C overnight. The following morning the diluted memapsin 2 solution was maintained at room temperature for 4-6 hours followed by adjusting pH to 8.5 and replacing the flasks to the 4 °C room. The same procedure was followed the next day with adjustment of pH to 8.0.

This memapsin 2 solution was allowed to stand at 4 °C for 2-3 weeks. The total volume of approximately 16 liters was concentrated to 40 mls using ultra-filtration (Millipore) and stir-cells (Amicon), and centrifuged at 140,000 xg at 30 minutes in a rotor pre-equilbrated to 4 °C. The recovered supernatant was applied to a 2.5 x 100 cm column of S-300 equilibrated in 0.4 M urea, 20 mM Tris-HCl, pH 8.0, and eluted with the same buffer at 30 ml/hour. The active fraction of memapsin 2 was pooled and further purified in a FPLC using a 1 ml Resource-Q (Pharmacia) column. Sample was filtered, and applied to the Resource-Q column equilibrated in 0.4 M urea, 50 mM Tris-HCl, pH 8.0. Sample was eluted with a gradient of 0 - 1 M NaCl in the same buffer, over 30 ml at 2 ml/min. The eluents containing memapsin 2 appeared near 0.4 M NaCl which was pooled for crystallization procedure at a concentration near 5 mg/ml.

The amino-terminal sequence of the protein before crystallization showed two sequences starting respectively at residues 28p and 30p. Apparently, the pro peptide of recombinant pro-memapsin 2 had been cleaved during the preparation by a yet unidentified proteolytic activity.

The activation of the folded pro-enzyme to mature enzyme, memapsin 2, was carried out as described above, i.e., incubation in 0.1 M sodium acetate pH 4.0 for 16 hours at 22 °C. Activated enzyme was further purified using anion-exchange column chromatography on Resource-Q anion exchange column. The

purity of the enzyme was demonstrated by SDS-gel electrophoresis. At each step of the purification, the specific activity of the enzyme was assayed as described above to ensure the activity of the enzyme.

Preliminary Crystallization with OM99-2

Crystal trials were performed on purified memapsin 2 in complex with a substrate based transition-state inhibitor OM99-2 with a Ki = 10 nM. OM99-2 is equivalent to eight amino-acid residues (including subsites S4, S3, S2, S1 S1', S2', S3' and S4' in a sequence EVNLAAEF) with the substitution of the peptide bond between the S1 and S1' (L-A) by a transition-state isostere hydroxyethylene. Purified M2 was concentrated and mixed with 10 fold excessive molar amount of inhibitor. The mixture was incubated at room temperature for 2-3 hours to optimize the inhibitor binding. The crystallization trial was conducted at 20 °C using the hanging drop vapor diffusion procedure. A systemic search with various crystallization conditions was conducted to find the optimum crystallization conditions for memapsin 2/OM99-2 inhibitor complex. For the first step, a coarse screen aimed at covering a wide range of potential conditions were carried out using the Sparse Matrix Crystallization Screen Kits purchased from Hampton Research. Protein concentration and temperature were used as additional variables. Conditions giving promising (micro) crystals were subsequently used as starting points for optimization, using fine grids of pH, precipitants concentration etc.

Crystals of memapsin-inhibitor complex were obtained at 30% PEG 8000, 0.1 M NaCocadylate, pH 6.4. SDS gel electrophoresis of a dissolved crystal verified that the content of the crystal to be memapsin 2. Several single crystals (with the sizes about 0.3 mm x 0.2 mm x 0.1 mm) were carefully removed from the cluster for data collection on a Raxis IV image plate. These results showed that the crystals diffract to 2.6 Å. ~~A typical protein diffraction pattern is shown in Figure 6.~~ An X-ray image visualization and integration software–Denzo, was used to visualize and index the diffraction data. Denzo identified that the primitive orthorhombic lattice has the highest symmetry with a significantly low distortion index. The unit cell parameters were determined

as: a=89.1 Å, b=96.6 Å, c=134.1 Å, $\alpha=\beta=\gamma=90^{o}$. There are two memapsin 2/OM99-2 complexes per crystallographic asymmetric unit, the V_m of the crystal is 2.9 $Å^3$/Da. Diffraction extinctions suggested that the space group is $P2_12_12_1$.

With diffraction of the current crystal to 2.6 Å, the crystal structure obtained from these data has the potential to reach atomic solution, i.e., the three-dimensional positions of atoms and chemical bonds in the inhibitor and in memapsin 2 can be deduced. Since memapsin 2 sequence is homologous with other mammalian aspartic proteases, e.g., pepsin or cathepsin D, it is predicted that the three dimensional structures of memapsin 2 will be similar (but not identical) to their structures. Therefore, in the determination of x-ray structure from the diffraction data obtained from the current crystal, it is likely the solution of the phase can be obtained from the molecular replacement method using the known crystal structure of aspartic proteases as the search model.

Further Crystallization Studies

Concentrated memapsin 2 was mixed with 10-fold molar excessive of the inhibitor. The mixture was incubated at room temperature for 2-3 hours to optimize inhibitor binding, and then clarified with a 0.2 micron filter using centrifugation. Crystals of memapsin 2-inhibitor complex were grown at 20 °C by hanging drop vapor diffusion method using equal volumes of enzyme-inhibitor and well solution. Crystals of quality suitable for diffraction studies were obtained in two weeks in 0.1 M sodium cacodylate, pH 7.4, 0.2 M $(NH_4)_2SO_4$, and 22.5% PEG8000. The typical size of the crystals was about 0.4 x 0.4 x 0.2 mm^3.

Diffraction data were measured on a Raxis-IV image plate with a Rigaku X-ray generator, processed with the HKL program package [Z. Otwinowski, W. Minor, Methods Enzymol. 276, 307 (1997)] A single crystal of approximately 0.4 x 0.4 x 0.2 mm^3 in size was treated with a cryo-protection solution of 25% PEG8000, 20% glycerol, 0.1 M sodium-cacodylate pH 6.6, and 0.2 M $(NH_4)_2SO_4$, and then flash-cooled with liquid nitrogen to about –180 °C for data collection. Diffraction was observed to at least 1.9 Å. The crystal form

belongs to space group $P2_1$ with two memapsin 2/OM99-2 complexes per crystallographic asymmetric unit and 56% solvent content.

Molecular replacement was performed with data in the range of 15.0-3.5 Å using program AmoRe, CCP4 package [Navaza, J., Acta Crystallog. Sect. A. 50, 157 (1994)]. Pepsin, a human aspartic protease with 22% sequence identity, was used as the search model(PDB id 1psn). Rotation and translation search, followed by rigid body refinement, identified a top solution and positioned both molecules in the asymmetric unit. The initial solution had a correlation coefficient of 22% and an R-factor of 0.51. The refinement was carried out using the program CNS [Brunger et al., Acta Crystallogr. Sect. D, 54, 905 (1998)]. 10% of reflections were randomly selected prior to refinement for R_{free} monitoring [Bruger, A.T., X-PLOR Version 3.1: A system for X-ray Crystallography and NMR, Yale University Press, New Haven, CT (1992)]. Molecular graphics program [Jones, T.A., et al., Improved methods for building protein models in electron denisty maps and location of errors in these models. Acta Crystallogr. Sect. A 47, 110 (1991)] was used for map display and model building. From the initial pepsin model, corresponding amino acid residues were changed to that of memapsin 2 according to sequence alignment. The side chain conformations were decided by the initial electron density map and a rotomer library. This model was refined using molecular dynamics and energy minimization function of CNS [Bruger, A.T., et al., Acta Crystallogr. Sect. D, 54, 905 (1998)]. The first cycle of refinement dropped the $R_{working}$ to 41% and the R_{free} to 45%. At this stage, electron densities in the omit map clearly showed the inhibitor configuration in the active site cleft. Structural features unique to memapsin 2 in chain tracing, secondary structure, insertions, deletions and extensions (as compared to the search model) are identified and constructed in subsequent iterations of crystallographic refinement and map fitting. The inhibitor was built into the corresponding electron density.

About 440 solvent molecules were then gradually added to the structure as identified in the |Fo|-|Fc| map contoured at the 3 sigma level. Non-crystallographic symmetry restriction and averaging were used in early stages

of refinement and model building. Bulk solvent and anisotropic over-all B factor corrections were applied through the refinement. The final structure was validated by the program PROCHECK Laskowski, R.A. et al., J. Appl. Crystallog. 26, 283 (1993) which showed that 95% of the residues are located in the most favored region of the Ramachandran plot. All the main chain and side chain parameters are within or better than the standard criteria. The final $R_{working}$ and R_{free} are 18% and 22% respectively. Refinement statistics are listed in Table 2.

Table 2. Data Collection and Refinement Statistics

A. Data Statistics	
Space group	$P2_1$
Unit Cell (a, b, and c in)	53.7, 85.9, 109.2
(, , and in degrees)	90.0, 101.4, 90.0
Resolution ( )	25.0-1.9
Number of observed reflections	144,164
Number of unique reflections	69,056
R_{merge}[a]	0.061 (0.25)
Data completeness (%) (25.0-1.9)	90.0 (68.5)
<I/ (I)>	13.7 (3.0)
B. Refinement Statistics	
$R_{working}$[b]	0.186
R_{free}[b]	0.228
RMS deviation from ideal values	
Bond length ( )	0.014
Bond angle (Deg)	1.7
Number of water molecules	445
Average B-factor (^{2})	
Protein	28.5
Solvent	32.2

[a] $R_{merge} = \sum_{hkl} \sum_{i} | I_{hkl,i} - \langle I_{hkl} \rangle | / \sum_{hkl} \langle I_{hkl} \rangle$, where $I_{hkl,i}$ is the intensity of the ith measurement and $\langle I_{hkl} \rangle$ is the weighted mean of all measurements of I_{hkl}.

[b] $R_{working\ (free)} = \sum | F_o| - | F_c| / \sum | F_o|$, where F_o and F_c are the observed and calculated structure factors. Numbers in parentheses are the corresponding numbers for the highest resolution shell (2.00-1.9 Å). Reflections with $F_o/ (F_o) >= 0.0$ are included in the refinement and R factor calculation.